Journal of Computational Physics 231 (2012) 2682–2690

Contents lists available at SciVerse ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Short note

A simple filter for detecting low-rank submatrices

Aaditya V. Rangan [1]

Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, NY 10012, United States

ARTICLE INFO

Article history:
Received 1 August 2011
Received in revised form 6 December 2011
Accepted 23 December 2011
Available online 4 January 2012

Keywords:
Random projection
Biclustering

ABSTRACT

We present a simple algorithm for detecting low-rank submatrices from within a much larger matrix. This algorithm relies on a basic geometric property of high-dimensional space: random 2-d projections of eccentric gaussian distributions are typically concentrated in opposite quadrants of the plane.

1. Introduction

Many techniques for data-analysis and matrix-compression take advantage of the reduction of dimensionality which can be attained if a submatrix within a larger matrix has low numerical-rank [1–10]. A natural question is: given a large matrix, how can one quickly detect submatrices with low numerical-rank? In this paper we present a very simple algorithm for detecting the largest submatrix C of low numerical-rank $k \sim 1-5$ within a larger matrix A which is given as data.

The core of the algorithm itself is very simple: Given a large $n \times m$ matrix A, we first form the 'binary' matrix B by sending each entry of a A to either +1 or -1, depending on its sign. Then we form $Z^{\text{row}} \in \mathbb{R}^n$ by taking the diagonal entries of $BB^{\mathsf{T}}BB^{\mathsf{T}}$, and we form $Z^{\text{ col}} \in \mathbb{R}^m$ by taking the diagonal entries of $B^{\mathsf{T}}BB^{\mathsf{T}}B$. We then eliminate the rows and columns of A for which Z^{row} and Z^{col} are small, and repeat this entire process. Eventually, after repeating this process multiple times, we will eliminate almost all the rows and columns of A, retaining only those rows and columns which form the low-rank submatrix C.

This algorithm makes use of the following geometric feature of high dimensional space: a random planar projection of an eccentric gaussian distribution is typically concentrated in non-adjacent quadrants. This fact implies that 2×2 submatrices of B (referred to as 'loops' in the following sections) contain substantial information about C, and that rows and columns of C will correspond to large values of Z^{row} and Z^{col}.

There are other approaches to finding certain kinds of low-rank submatrices, such as nuclear norm minimization [11] and adaptive dissection of projective space [12]. The advantages of the approach presented in this paper include:

- Ease of implementation: this entire algorithm can be implemented with a loop comprising two matrix multiplications, requiring only a few lines of Matlab code.
- Reasonable efficiency: this algorithm requires only a handful of matrix multiplications, incurring a complexity of $O(mn\min(m,n))$ along with a low constant factor.
- Flexibility: this algorithm can be easily tailored to a variety of applications in data-analysis.

[1] Supported by NSF Grant DMS-0914827.
E-mail address: rangan@cims.nyu.edu
URL: http://www.cims.nyu.edu/~rangan

URLs: http://www.cims.nyu.edu/~rangan

0021-9991/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2011.12.032

An inevitable disadvantage of the approach presented in this paper is that it is not always guaranteed to work; the underlying submatrix detection problem is NP-hard [13,14]. However, under rather general conditions this scheme will detect any sufficiently large C with high probability. Importantly, this algorithm still often succeeds even when C is very small relative to A (e.g., of size $\sim \sqrt{n} \times \sqrt{m}$).

In the remainder of this paper we provide a justification and full description of this algorithm, as well as examples.

2. Detecting a low-rank submatrix

Generally speaking, there is one basic problem we will consider. This problem involves finding the largest submatrix of low numerical-rank from within a larger matrix. This problem (stated formally below) is often called the 'biclustering' problem in data-analysis [15,16].

Problem 1. Assume that we are given an $n \times m$ matrix A, and that an $n_C \times m_C$ submatrix C of A has low numerical-rank k (i.e., $k \leqslant 5$). Assuming that C is the largest such submatrix within A, is it possible to find C quickly?

First we will discuss a geometric feature of high-dimensional space, and then we will discuss how to use this feature to solve Problem 1.

2.1. Planar projections of eccentric gaussian distributions are concentrated in non-adjacent quadrants

First let us define the 'binarization operator'.

Definition 2. The operator $\mathbb{B}[\cdot]$ replaces each entry of its input with either +1 or -1, depending on its sign.

So, for example, $\mathbb{B}([3.1, -0.5]) = [+1, -1]$.

Now let us define what we mean by 'numerical-rank':

Definition 3. A matrix is of numerical-rank k with 'error' ε if the $(k+1)$st singular-value σ_{k+1} of this matrix is equal to $\varepsilon\sigma_k$, where σ_k is the kth singular-value of the matrix.

Definition 4. A gaussian-distribution ρ on $\mathbb{R}^m$ is of numerical-rank k with error ε if the $(k+1)$st principal-value of ρ is equal to ε times the kth principal-value of ρ.

The error ε is a measure of how well the matrix or distribution under consideration can be approximated within a k-dimensional subspace. If $\varepsilon = 0$, then the matrix or distribution is exactly rank-k.

One simple fact about distributions with low numerical-rank is that, when projected onto a randomly oriented 2-dimensional plane, most of the mass of these distributions lies within non-adjacent quadrants. To make this statement more precise, let us assume that the gaussian distribution ρ is a randomly oriented ε-error rank-k distribution on $\mathbb{R}^m$ with k principal-values equal to 1, and $m-k$ principal-values equal to ε. Let us assume that $P^{2\leftarrow m} : \mathbb{R}^m \to \mathbb{R}^2$ is an orthogonal projection onto a plane, such as 2 arbitrary rows or columns of the $m \times m$ identity-matrix. The distribution $\tilde{\rho} = P^{2\leftarrow m}\rho$ is a distribution on $\mathbb{R}^2$. Let v_1 and v_2 be two vectors drawn independently from $\tilde{\rho}$. Let us define $g_{\varepsilon,k,m}$ to be the probability that v_1 and v_2 lie in non-adjacent quadrants of $\mathbb{R}^2$. This probability $g_{\varepsilon,k,m}$ can also be thought of as the probability that $\mathbb{B}(v_1) \| \mathbb{B}(v_2)$.

We can estimate $g_{\varepsilon,k,m}$, and state the following.

Claim 5. *$g_{\varepsilon,k,m}$ is substantially greater than 1/2 when k is not too large, and $\varepsilon < 1/\sqrt{m}$.*

This claim is illustrated in Fig. 1, which shows plots of $g_{\varepsilon,k,m}$ for $k = 1, \ldots, 6$. For small fixed k and large m the value $g_{\varepsilon,k,m}$ is essentially determined by the product $\varepsilon\sqrt{m}$, and is significantly greater than 1/2 as long as $\varepsilon\sqrt{m} \lesssim 1$ (as discussed further in Appendix A.2). The Claim 5 has many useful ramifications, some of which we will discuss later in Section 2.2. Note that, when $k = 1$ and $\varepsilon = 0$, the distribution ρ is a line, and since any planar projection of a line is still a line, the probability $g_{0,1,m} = 1$. If $k \gg 1$ and ε is close to 1, then the distribution ρ is nearly spherical, and any planar projection of ρ will be roughly uniformly distributed, implying that $g_{k,\varepsilon,m} \sim 1/2$. Importantly however, as long as k and ε are both small, then $g_{\varepsilon,k,m}$ is significantly greater than 1/2.

Note also that, since ρ is randomly oriented, $g_{\varepsilon,k,m}$ can be thought of as the probability that any 2×2 submatrix of $\mathbb{B}([v_1, v_2])$ is rank-1, rather than rank-2. Since we will refer to these types of submatrices frequently, we will refer to a 2×2 submatrix as a 'loop'.

2.2. Interpretations of Claim 5

Claim 5 can be used to justify the algorithm presented later in this paper (see Section 3). The basic observation is that, when considering Problem 1, a loop (i.e., a 2×2 submatrix) of $\mathbb{B}(C)$ is more likely to be rank-1 (and less likely to be rank-2) than a loop of $\mathbb{B}(A)$ is.